CONTENT BASED VIDEO RETRIEVAL

P Saranya, G Sai Kumar Reddy, T Sharuthma, L Sukeshwar Reddy, B Saitha, G Rehan.

Department of Electronics and Communication Engineering, Siddharth Institute of Engineering and Technology, Puttur, Andhra Pradesh, India.

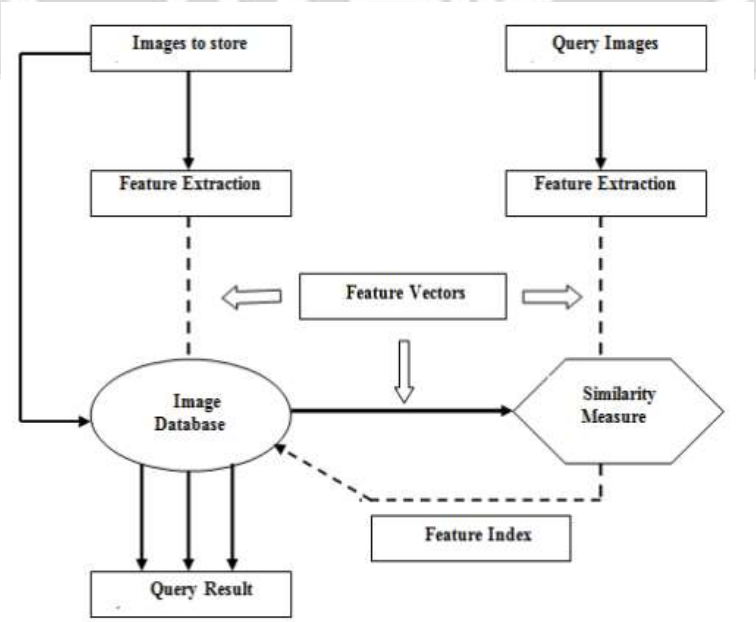
Abstract

Content-based retrieval allows finding information by searching its content rather than its attributes. The challenge facing content-based video retrieval (CBVR) is to design systems that can accurately and automatically process large amounts of heterogeneous videos. Moreover, content-based video retrieval system requires in its first stage to extract the video into separate frames. Afterwards features are extracted for video frames. And finally, choose a similarity/classifier metric and by machine learning algorithm that is efficient enough to retrieve query – related videos results. Histogram of Oriented Gradients (HOG) features are extracted for video frame and Random Forest Classifier, a machine learning algorithm is used for classification of video frame.

Keywords: *Content Based Video Retrieval (CBVR), Histogram of Oriented Gradients, Machine Learning, Random Forest Classifier.*

1. INTRODUCTION

An image retrieval is retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, title or descriptions to the image so that retrieval can be performed over the annotation words. Manual video annotation is time-consuming, laborious and expensive to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.



Search methods:

Video search is a specialized data search used to find videos. To search for videos, a user may provide query terms such as keyword, video file/link, or click on some video, and the system will return videos "similar" to the query. The similarity used for search criteria could be Meta tags, color distribution in videos, region/shape attributes, etc.

- Video Meta search - search of videos based on associated metadata such as keywords, text, etc.
- Content-based video retrieval – the application of computer vision to the video retrieval. CBVR aims at avoiding the use of textual descriptions and instead retrieves videos based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query video or user-specified video features.
- List of CBVR Engines - list of engines which search for videos-based video visual content such as color, texture, shape/object, etc.
- Video collection exploration - search of videos based on the use of novel exploration paradigms.

Video Meta search:

Video Meta search (or video search engine) is a type of search engine specialized on finding pictures, videos, animations etc. Like the text search, video search is an information retrieval system designed to help to find information on the Internet and it allows the user to look for videos etc. using keywords or search phrases and to receive a set of thumbnail videos, sorted by relevancy. Specialized search engines, like in the fields of video search, are among the fastest growing search services on the internet. In 2005 alone the number of video searches increased by 91%. The most common search engines today offer video search such as Google, Yahoo or Bing!

A search engine is a software system that is designed to carry out web searches (Internet searches), which means to search the World Wide Web in a systematic way for particular information specified in a textual web search query. The search results are generally presented in a line of results, often referred to as search engine results pages (SERPs) The information may be a mix of links to web pages, videos, videos, info graphics, articles, research papers, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler. Internet content that is not capable of being searched by a web search engine is generally described as the deep web.
Metadata is "data that provides information about other data". In other words, it is "data about data". Many distinct types of metadata exist, including descriptive metadata, structural metadata, administrative metadata, reference metadata and statistical metadata.

- Descriptive metadata is descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
- Structural metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.
- Administrative metadata is information to help manage a resource, like resource type, permissions, and when and how it was created.
- Reference metadata is information about the contents and quality of statistical data.

Statistical metadata, also called process data, may describe processes that collect, process, or produce statistical data.

How video search works?

A common misunderstanding when it comes to video search is that the technology is based on detecting information in the video itself. But most video search works as other search engines. The metadata of the video is indexed and stored in a large database and when a search query is performed the video search engine looks up the index, and queries are matched with the stored information. The results are presented in order of relevancy. The usefulness of a video search engine depends on the relevance of the results it returns, and the ranking algorithms are one of the keys to becoming a big player.

Some search engines can automatically identify a limited range of visual content, e.g. faces, trees, sky, buildings, flowers, colors etc. This can be used alone, as in content-based video retrieval, or to augment metadata in a video search. When performing a search the user receives a set of thumbnail videos, sorted by relevancy. Each thumbnail is a link back to the original web site where that video is located. Using an advanced search option, the user can typically adjust the search criteria to fit their own needs, choosing to search only videos or animations, color or black and white, and setting preferences on video size.

Video search providers:

- AltaVista
- Corbis
- GazoPa (similar video search, has been shut down for consumer, still available for business user)
- Google Video Search (also reverse video search)
- Live Search from Microsoft
- Macroglossa (visual search engine)
- Picollator
- Picsearch
- Pixsta
- TinEye (only reverse video search)
- Yandex Search (also reverse video search)

Video collection:

Video collection exploration is a mechanism to explore large digital video repositories. The huge number of digital videos produced every day through different devices such as mobile phones bring forth challenges for the storage, indexing and access to these repositories. Content-based video retrieval (CBVR) has been the traditional paradigm to index and retrieve videos. However, this paradigm suffers of the well-known semantic gap problem. Video collection exploration consists of a set of computational methods to represent, summarize, visualize and navigate video repositories in an efficient, effective and intuitive way.

Automatic summarization consists in finding a set of videos from a larger video collection that represents such collection. Different methods based on clustering have been proposed to select these video prototypes (summary). The summarization process addresses the problem of selecting a representative set of videos of a search query or in some cases, the overview of a video collection.

Automatic summarization is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content. In addition to text, videos and videos can also be summarized. Text summarization finds the most informative sentences in a document, video summarization finds the most representative videos within a video collection [citation needed] video summarization extracts the most important frames from the video content.

There are two general approaches to automatic summarization: extraction and abstraction.
Extraction-based summarization:
Here, content is extracted from the original data, but the extracted content is not modified in any way. Examples of extracted content include key-phrases that can be used to "tag" or index a text document, or key sentences (including headings) that collectively comprise an abstract, and representative videos or video segments, as stated above. For text, extraction is analogous to the process of skimming, where the summary (if available), headings and subheadings, figures, the first and last paragraphs of a section, and optionally the first and last sentences in a paragraph are read before one chooses to read the entire document in detail. Other examples of extraction that include key sequences of text in terms of clinical relevance (including patient/problem, intervention, and outcome).

Abstraction-based summarization:
This has been applied mainly for text. Abstractive methods build an internal semantic representation of the original content, and then use this representation to create a summary that is closer to what a human might express. Abstraction may transform the extracted content by paraphrasing sections of the source document, to condense a text more strongly than extraction. Such transformation, however, is computationally much more challenging than extraction, involving both natural language processing and often a deep understanding of the domain of the original text in cases where the original document relates to a special field of knowledge. "Paraphrasing" is even more difficult to apply to video and video, which is why most summarization systems are extractive.

Aided summarization:
Approaches aimed at higher summarization quality rely on combined software and human effort. In Machine Aided Human Summarization, extractive techniques highlight candidate passages for inclusion (to which the human adds or removes text). In Human Aided Machine Summarization, a human post-processes software output, in the same way that one edits the output of automatic translation by Google Translate.

Video collection visualization is the process of visualize a set of videos using a visualization metaphor, in which a video similarity function is used to represent video relations in a visualization layout. Information visualization is an active area that investigates new ways to visualize information by using visualization metaphors. Particularly, new ways of visualizing video collections are being investigated, which propose conventional and unconventional visualization metaphors. If the videos are sorted according to their similarities, a hierarchic video browsing approach similar to cartographic services such as Google Maps can be used. Picsbuffet is an online demo of such an approach.

Information visualization or information visualization is the study of (interactive) visual representations of abstract data to reinforce human cognition. The abstract data include both numerical and non-numerical data, such as text and geographic information. The naming of subfields is sometimes confusing. One accepted definition is that it's information visualization when the spatial representation is chosen, whereas it's scientific visualization when the spatial representation is given.

The field of information visualization has emerged "from research in human-computer interaction, computer science, graphics, visual design, psychology, and business methods. It is increasingly applied as a critical component in scientific research, digital libraries, data mining, financial data analysis, market studies, manufacturing production control, and drug discovery.
Information visualization presumes that "visual representations and interaction techniques take advantage of the human eye’s broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once. Information visualization focused on the creation of approaches for conveying abstract information in intuitive ways."

Video collection interaction consists in offering users mechanisms to feedback video search systems. In this interaction process, the system learns from user feedback to retrieve results more precise and relevant to the user.

Data Scope:

It is crucial to understand the scope and nature of video data in order to determine the complexity of video search system design. The design is also largely influenced by factors such as the diversity of user-base and expected user traffic for a search system. Along this dimension, search data can be classified into the following categories:

- Archives - usually contain large volumes of structured or semi-structured homogeneous data pertaining to specific topics.
- Domain-Specific Collection - this is a homogeneous collection providing access to controlled users with very specific objectives. Examples of such a collection are biomedical and satellite video databases.
- Enterprise Collection - a heterogeneous collection of videos that is accessible to users within an organization’s intranet. Pictures may be stored in many different locations.
- Personal Collection - usually consists of a largely homogeneous collection and is generally small in size, accessible primarily to its owner, and usually stored on a local storage media.
- Web - World Wide Web videos are accessible to everyone with an Internet connection. These video collections are semi-structured, non-homogeneous and massive in volume, and are usually stored in large disk arrays.

2.EXISTING METHOD:

Content Based Image Retrieval (CBIR) is performed. CBIR is a technique used for automatic retrieval of images in a large database that perfectly matches the query image. It is difficult to search the desired image stored in a huge database. Thus, to automatically search the desired image from large database, there is a need to develop an expert technique. Basically, two approaches are common in practice to retrieve the image in database: text based and visual based. Generally, the images are searched on the basis of text by using Google, Yahoo etc., i.e., the images stored in text annotation and user types in a sequence of text. Thus, the images are retrieved on the basis of these texts.

To increase the accuracy and efficiency of image retrieval system the Content Based Image Retrieval was introduced in 1990. Content Based Image Retrieval technique used visual content of query image perfectly, matched it with the database images and measures similarity based on color, texture and shape. On the basis of similarity, the images were retrieved

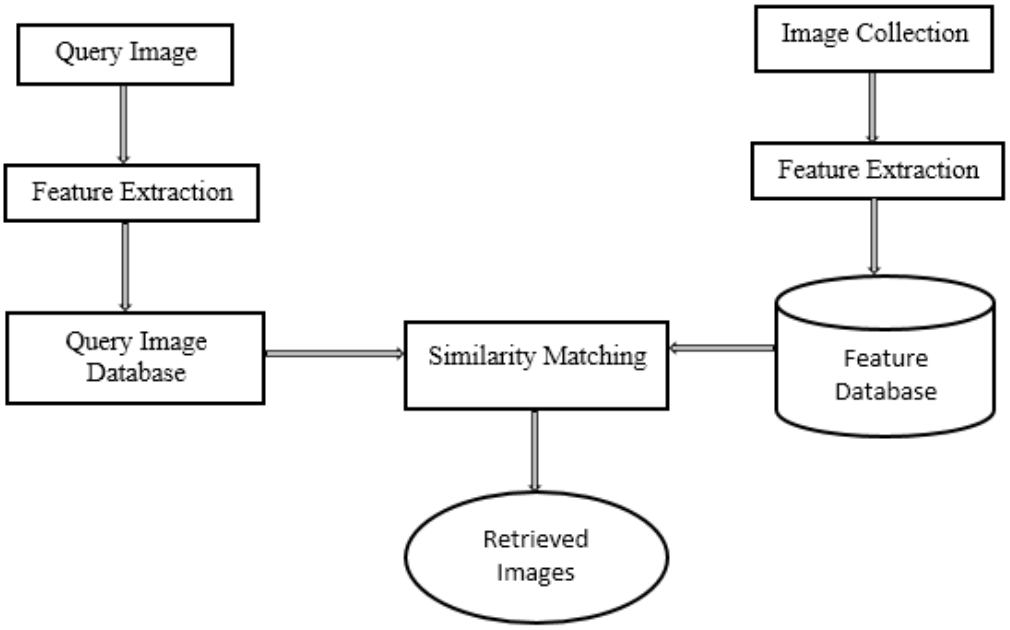


Fig.2 Block diagram of Existing method

Content Based Image Retrieval (CBIR):

The query input will be image. LBP features are extracted from image. Random forest classifier is used for classification of query image. Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

A. Image Retrieval by Color:
Color feature is a sensitive and understandable feature of the image, and normally histogram techniques are used to demonstrate it. The advantage of Color histogram technique is that, it has high speed, it does not require huge memory and not susceptible with the change in images size and other parameters. On the basis of Hue Saturation Value of color space, the color feature vector of query image and database image is calculated. To increase the accuracy of image retrieval, histogram technique is treated to be superior to Hue Saturation Value. Following steps are involved in retrieval of image by using color features.

1. Feature Extraction
2. Histogram calculation
3. Similarity Matrix calculation
4. Dissimilarity calculation
5. Arrangement of Images in ascending order.

B. Image Retrieval by Texture

Texture is used to illustrate a basic constituent of images which are arranged uniformly. The region of an image can be determined by texture segmentation. After finding the region of image, their bounding boxes may be used to retrieve the formation like an R-tree. The problem occurring with dimension and cross correlation also affects the texture and it can be solved by comparable technique used in color retrieval. Periodicity and scale are the merits of texture. Texture has qualities like periodicity and scale; it can be expressed in terms of Contrast, direction and thickness. The natural possessions of surface is texture and it demonstrate visual pattern, it include significant information associated to the structural arrangement of the surface like buildings, sea, plants, textile. Accessing the related image by using texture technique, two most important issues involved namely arithmetic analysis and structural analysis. When the texture component can be clearly identified, structural analysis is used whereas for micro texture arithmetic analysis is used.

Arithmetic process distinguishes dissimilarity of power in a texture window. For example, process contains contrast (high contrast tiger skin vs. low contrast monkey skin), coarseness (dense stones vs. coarse gravel), and directionality (directed cloth vs. undirected grass). Texture element in the image obtained from structural technique, identifies their shapes and calculates their placement policies. This set of laws are used to illustrate how texture element are to be found in image and determined the number of instantaneous neighbors, number of unit space element and whether they are laid out equally or not.

C. Image Retrieval by Shape

Next important method for retrieving image is by using its shape feature. Shape representations can be usually divided into two categories, Region based and Boundary based. Simply the external edge of the shape is used in Boundary based representation. In this method, the outer description of the region, like the pixels near the object edge is measured. However, another technique is completely different from previous method. Region based method uses the complete shape area of image by relating the considered area via its inner quality; i.e., the pixels enclosed in that area.

Texture:

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image the texture is located. Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated. The problem is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as silky, or rough. Other methods of classifying textures include:

- Co-occurrence matrix
- Laws texture energy
- Wavelet transform
- Orthogonal transforms (Discrete Tchebichef moments)

Shape:

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined first applying segmentation or edge detection to an image. Other methods use shape filters to identify given shapes of an image. Shape descriptors may also need to be invariant to translation, rotation, and scale. Some shape descriptors include:

- Fourier transform
- Moment invariant

Disadvantages:

- One of disadvantage of LBP is that many various structural patterns may have the identical LBP
- The LBP feature has the drawback that it is not too robust on flat image areas. An image region is said to be "flat" if it has a nearly uniform intensity.
- Not applicable for video retrieval

PROPOSED METHOD

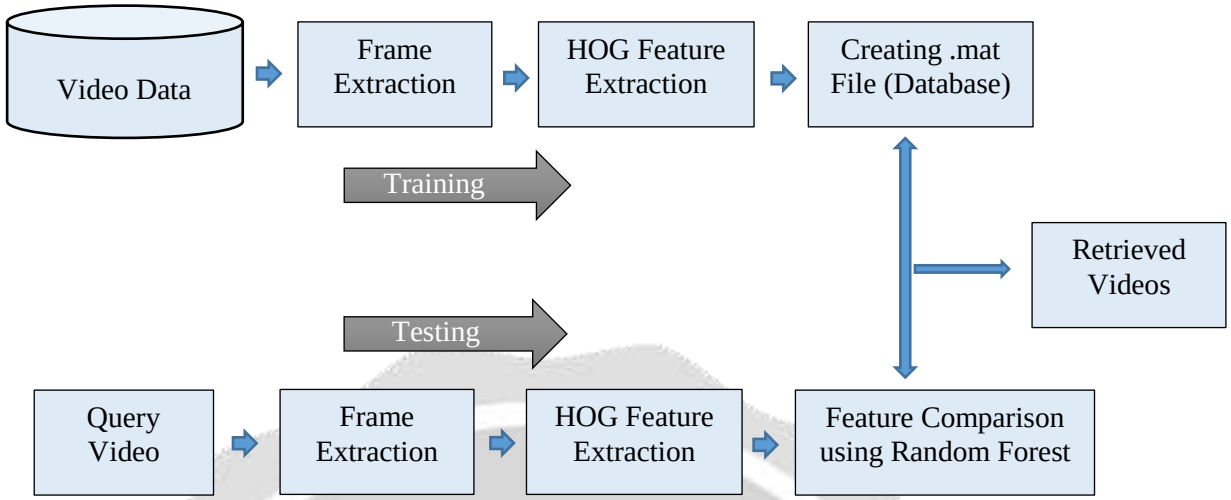


Fig: Block Diagram of Proposed Method

- Video data is collected from Online sources like YouTube
- In video and animation, frames are individual pictures in a sequence of images. For example, a Flash movie you see on the Web may play 12 frames per second, creating the appearance of motion. Most video is shot at 24 or 30 frames per second, or FPS.
- HOG, or Histogram of Oriented Gradients, is a feature descriptor that is often used to extract features from image data. It is widely used in computer vision tasks for object detection.
- The HOG descriptor focuses on the structure or the shape of an object. Now you might ask, how is this different from the edge features we extract for images? In the case of edge features, we only identify if the pixel is an edge or not. HOG is able to provide the edge direction as well. This is done by extracting the gradient and orientation (or you can say magnitude and direction) of the edges.
- Additionally, these orientations are calculated in 'localized' portions. This means that the complete image
- Files with a .mat extension contain MATLAB formatted data, and data can be loaded from or written to these files using the functions load and save, respectively.
- Video querying refers to the problem of finding objects that are relevant to a user query within video databases
- A video retrieval system is a computer system for browsing, searching and retrieving videos from a large database of digital videos

2. EXPERIMENTAL RESULTS

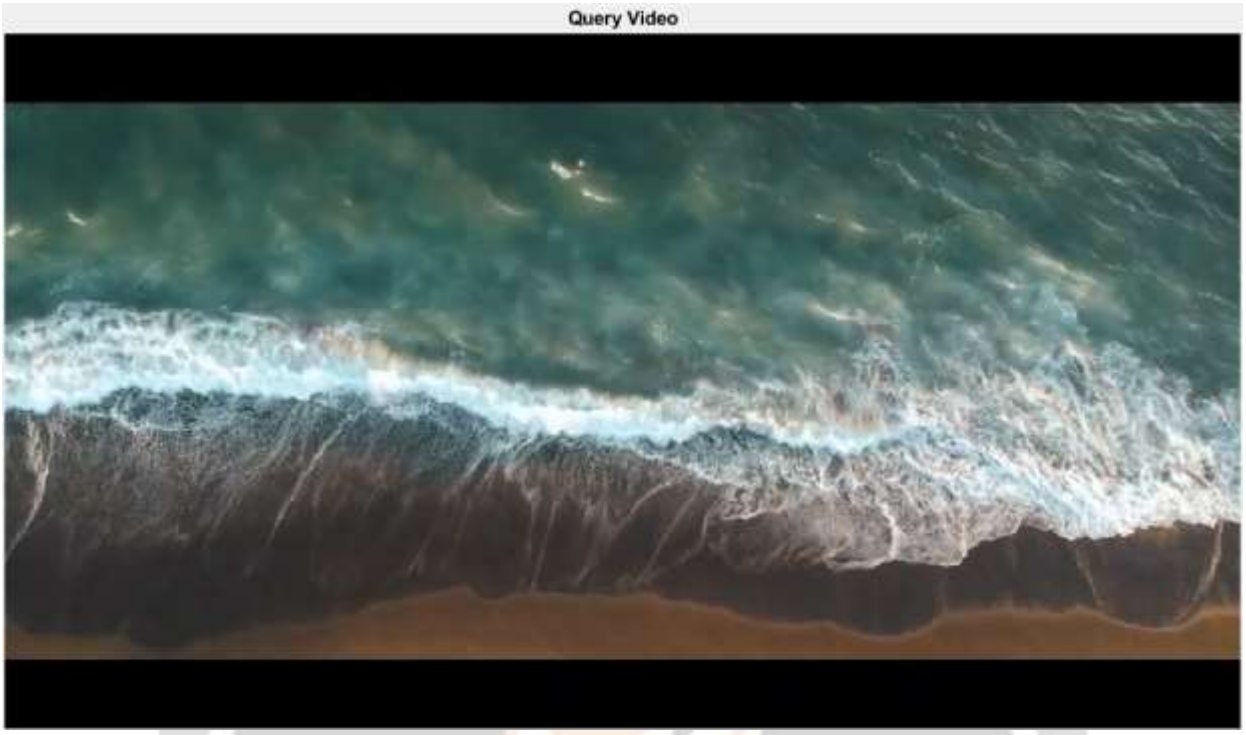


Figure 2 : Query Video

Fig: Reterived Videos

The proposed system uses multiple features like edge detection features and texture features. The Canny Edge Detection is used for edge detection from the image and the GLCM texture feature is used to extract texture feature from the image. The work done considering a combination of GLCM Texture Feature and Color Histogram showed an accuracy of only about 50-55%, because the color histogram is a low-level feature, sensitive to noise. Whereas, retrieving videos using a combination of edge detection and the GLCM texture feature showed a much higher and better combination, it showed an accuracy of 65-70%. Hence, in the proposed work, a combination of Edge Detection and GLCM Feature Extraction has been used to retrieve relevant videos. The same process of feature extraction is repeated for the query image input by the user. Then the Euclidean distance is calculated to measure the similarity between the query image and the frames and hence retrieve the desired videos.

3. CONCLUSION & FUTURE SCOPE

. This paper has presented a new retrieval system for digital videos which are based on HOG features. Here, Content Based Video Retrieval (CBVR) has been implemented in Matlab Software. The proposed method has the supremacy of capturing both spatial and temporal features, and by making use of multiple features the retrieval performance is improved.

In the future, the retrieval of required content over large video dataset using deep learning methods makes easier to analyze the video for crime investigations. Conflict of Interest. The authors declare that there is no conflict of interest in this work.

REFERENCES:

[1] V.Kumar, V.Tripathi, and B.Pant, "Unsupervised learning of visual representation via Rotation and future frame prediction for video retrievel.(2021)
[2] J.Shao, X. Wen ,B.Zhao temporal context aggregation for video retrieval with contrastive learning.aug2020
[3] C. Tang and A.Arnold, neural document expansion for ad-hoc information retrieval dec2020.
[4] V. Nagarajan, "Noise Removal Through the Exploration of Subjective and Apparent Denoised Patches Using Discrete Wavelet Transform", IETE Journal of Research, ISSN: 0377-2063, 2019, DOI:10.1080/03772063.2019.1569483.
[5] X.Li, deep learning for video retrival by natural language, in proceedings of the first international work shop on fairness, accountability, oct2019
[6] R.Nogueira , W.Yang document expansion by query prediction, apr2019.
[7] P.Ambawane D.Bharatia and P.Rane , smart e-stick for visually impaired using video intelligence API in 2019 IEEE Bombay section signature conference .
[8] B. M¨unzer, M.J. Primus, M. Hudelist, C. Beecks, W. H¨urst, K. Schoeffmann "Content based Video Ret rieval and Human Computation unite**",** 978-1-5090-6067-2 © 2017 IEEE.
[9] Manisha Verma, Balasubramanian Raman **"**A Hierarchical Shot Boundary Detection Algorithm Using Global and Local Features", December 2017 .
[10] Tejeswi Potluri, Gnaneshwara O Nitta **"**Content based Video Retrieval using Dominant color of the truncated blocks of Frame", Journal of Theoretical and Applied Information Technology 2016.
[11] Janarthanan.Y, Balajee J.M and Srinivasa Raghavan.S **"**Content based Video Retrieval and Analysis using image processing**"**, International Journal of Pharmacy and Technology, December 2017.
[12] Priyanka Ramesh Malode, S. V. Gumaste **"**Feature Extraction from Video Using Ordered Dither Block Truncation Coding (ODBTC)And RGB Color Space Model ", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 6, June 2016.
[13] K.S. Thakre, A.M. Rajurkar and R.R. Manthalkar "Video partitioning and secured key frame extraction of MPEG Video**",** International Conference on Information Security & Privacy (ICISP2017), 11-12 December 2017, Nagpur, INDIA.
[14] Aasif Ansari and Muzammil H Mohammed "Content based Video Retrieval Systems-Methods, Techniques and Challenges", International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 7, February 2016.
[15] K. Sakthi dasan @ Sankaran, S. Bhuvaneshwari and Dr.V. Nagarajan "A new edge preserved technique using iterative median filter" in IEEE International Conference on Communication and Signal Processing(ICCSP 2014), pp:1750 – 1754, 2016.